

PATENT ABSTRACTS OF JAPAN

(11)Publication number :

2002-091472

(43)Date of publication of application : 27.03.2002

(51)Int.Cl.

G10L 13/00
G10L 21/06
G06F 17/30
G06F 17/60
G09B 19/00
G10L 11/04
G10L 11/02
G10L 13/08
G10L 15/00
G10L 15/02
// G09B 19/04
G09B 19/06

(21)Application number : 2000-283480

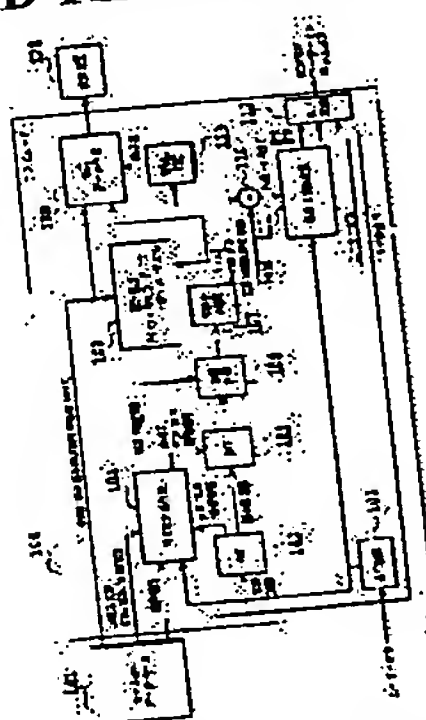
(71)Applicant : NIPPON HOSO KYOKAI <NHK>

(22)Date of filing :

19.09.2000

(72)Inventor : TSUGI TORU

(54) RHYTHM DISPLAY DEVICE, AND REPRODUCING DEVICE AND SIMILARITY
JUDGING DEVICE FOR VOICE LANGUAGE AND VOICE LANGUAGE PROCESSOR
AND RECORDING MEDIUM



(57)Abstract:

PROBLEM TO BE SOLVED: To visually and audibly support language learning or voice language research as for a rhythm display device, a reproducing device, and a similarity judging device for voice language, a voice language processor, and a recording medium.

SOLUTION: A model voice is fetched from a data base 101 for preliminarily analyzing and storing the characteristics of the rhythm of voice language, and a user voice whose contents are the same as those of the model voice is inputted, and the rhythm of an input voice corrected according to the result of the temporal correspondence of the model voice and the user voice is displayed.

The rhythm of the user voice is analyzed by an analyzing part 103. Then, the temporal correspondence of the analyzed result and the model voice is operated by a temporal corresponding part 104 and a correcting part 105. Display data to be graphically displayed are generated according to the corrected result by a time expanding part 106, a smoothing/interpolating part 107, a difference calculating part 109, and a display data generating part 108.

Detailed Description of the Invention:

....
[0046]

The processing by the computer 100 will be schematically described. First, when a user utters a voice (hereinafter referred to as "a user voice") to a microphone (not shown) connected to the computer 100 imitating a model voice, by listening to the model voice via the D/A converting part 112 and a speaker (not shown) or by referring to the time loci of the fundamental frequencies of the model voice which is displayed on the screen of the display device 120, the "user voice" is acquired into the computer 100 via the A/D converting part 102 as a voice waveform. The time loci of the fundamental frequencies of the acquired voice waveform are automatically displayed on the screen of the display device 120, and the user can visually know difference between the time loci of the fundamental frequencies of what he uttered and those of the model voice.

[0047]

At the same time, a converted voice which has been rhythm-converted is generated by the fundamental frequency converting part 111, the converted voice keeping the voice quality of the user voice but having the rhythm of the model voice. Therefore, the user can not only auditorily compare the user voice and the model voice by reproduced sounds from the speaker but also auditorily compare the voices before and after the rhythm change, which have the same user voice nature.

....
[0062]

Next, description will be made on the procedure for creation of fundamental-frequency-converted voice by the fundamental frequency converting part 111 or the like.

[0063]

First, a smoothed fundamental frequency time track $P_m(t)$ of the model voice is acquired from the model voice database part 101 by the difference calculating part 109, and difference ($p_m(t) - p_u(t)$) between the smoothed fundamental frequency time locus $P_m(t)$ and a smoothed fundamental frequency time locus $P_u(t)$ of the user voice by a smoothing/interpolation part 107 is determined. However, this difference includes difference between the average voice heights of both. Therefore, average values P_{mav} and P_{uav} for the whole phrases of $P_m(t)$ and $P_u(t)$, respectively, are determined by the difference calculating part 109. By subtracting the difference between them from the above difference to cancel, pure difference between temporal changes of the fundamental frequencies $P_d(t) = (P_m(t) - P_u(t)) - (P_{mav} - P_{uav})$ is calculated.

[0064]

Next, by adding the difference $P_d(t)$ calculated by the difference calculating part 109 and the fundamental frequency of the user voice before being smoothed $P_o(t)$ by the addition part 110, the temporal change of the fundamental frequency of the final converted voice $P_o(t) + P_d(t)$ is obtained. Thereby, synthesized voice is created in which the fundamental frequency during the voiced sound section of the user voice has replaced the fundamental frequency during the voiced sound section of the model voice associated with the voiced sound section of the user voice as a result of temporal association.

....
[0066]

That is, conversion of the fundamental frequency of the A/D converted user voice waveform (the whole time loci are shifted at the frequency axis so that the average value of the time track of the fundamental frequency which has replaced the fundamental frequency during the voiced sound section of the model voice may correspond to the average value of the time loci of the fundamental frequency of the user voice) and conversion of the time length of continuance of each sound are sequentially performed by the fundamental frequency converting part 111 (for example, in the method described on pp. 387-396 in The Transactions of the Institute of Electronics Information and Communication Engineers of Japan A, Vol. J73-An o.3) in accordance with the temporal change of the fundamental frequency $P_o(t) + P_d(t)$ by the addition part 110; converted voice in which the rhythm has replaced with that of the model voice is generated; and D/A converted voice is outputted from the speaker (not shown).

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号
特開2002-91472
(P2002-91472A)

(43)公開日 平成14年3月27日 (2002. 3. 27)

(51)Int.Cl. ⁷	識別記号	F I	テ-マ-コ-ト*(参考)
G 1 0 L 13/00		G 0 6 F 17/30	1 7 0 E 2 C 0 2 8
	21/06		3 5 0 C 5 B 0 4 9
G 0 6 F 17/30	1 7 0	17/60	1 2 8 5 B 0 7 5
	3 5 0	G 0 9 B 19/00	Z 5 D 0 1 5
17/60	1 2 8	19/04	5 D 0 4 5

審査請求 未請求 請求項の数26 O L (全 12 頁) 最終頁に続く

(21)出願番号	特願2000-283480(P2000-283480)	(71)出願人	000004352 日本放送協会 東京都渋谷区神南2丁目2番1号
(22)出願日	平成12年9月19日(2000. 9. 19)	(72)発明者	都木 徹 東京都世田谷区砦一丁目10番11号 日本放送協会 放送技術研究所内
		(74)代理人	100077481 弁理士 谷 義一 (外2名)

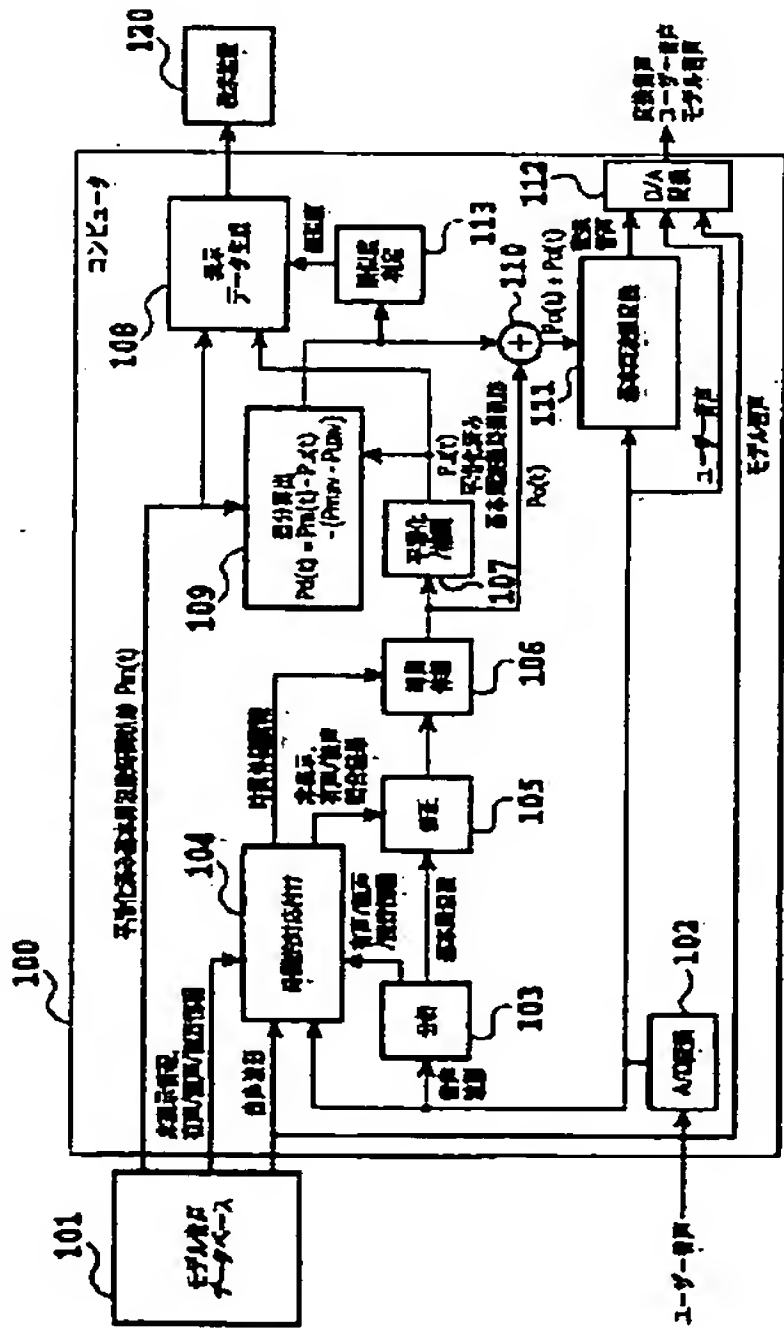
最終頁に続く

(54)【発明の名称】 音声言語の韻律表示装置、再生装置、類似度判定装置、音声言語処理装置、および記録媒体

(57)【要約】

【課題】 音声言語の韻律表示装置、再生装置、類似度判定装置、音声言語処理装置、および記録媒体に関し、語学学習や音声言語の研究を視覚的、聴覚的に支援する。

【解決手段】 本発明装置は、音声言語の韻律の特徴を予め分析して記憶したデータベース101からモデル音声を取り込むとともにモデル音声と同一内容のユーザー音声を入力し、モデル音声とユーザー音声の時間的対応付けの結果に従って修正された当該入力音声の韻律を表示する。分析部103により、ユーザー音声の韻律を分析する。時間的対応付け部104および修正部105により、この分析結果とモデル音声との時間的対応付けを行なう。時間伸縮部106、平滑化／補間部107、差分算出部109、および表示データ生成部108により、修正結果に従って、グラフ的に表示を行なうための表示データを生成する。



【特許請求の範囲】

【請求項1】 音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声の韻律を表示する音声言語の韻律表示装置であって、前記入力音声の韻律を分析する分析手段と、前記分析手段による分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正手段と、前記修正手段による修正結果に従って、前記表示をグラフィックに行なうための表示データを生成する表示手段とを備えたことを特徴とする韻律表示装置。

【請求項2】 請求項1において、前記修正手段は、前記入力音声および前記モデル音声それぞれの無音区間を省いてから時間的対応付けを行うことを特徴とする韻律表示装置。

【請求項3】 請求項1において、前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析手段を備え、前記表示手段は、前記修正結果に平滑化処理を行なうことで、前記基本周波数の時間軌跡を滑らかに変化させて前記表示を行なわせることを特徴とする韻律表示装置。

【請求項4】 請求項1において、前記表示手段は、前記修正手段による前記時間的対応付けの結果、前記モデル音声の有声部が前記分析結果の無声部に対応すると判定された場合、当該対応区間の基本周波数の時間軌跡を周辺の有声部の値を用いて補間し、前記モデル音声の無声部が前記分析結果の有声部に対応すると判定された場合、当該対応区間は非表示とすることを特徴とする韻律表示装置。

【請求項5】 請求項1において、前記表示手段は、前記分析結果の有声音区間の両端部で抽出された基本周波数が時間変化に対し不規則に変化すると前記修正手段によって判定された場合、当該変化部分のデータを前記有声音区間の別の部分の基本周波数の時間軌跡の値を用いて補間することを特徴とする韻律表示装置。

【請求項6】 請求項5において、前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析手段を備え、前記表示手段は、前記分析結果の所定の有声音区間の端部において、近傍の複数のフレームの基本周波数と前記有声音区間の長さとの前記有声音区間における時間位置によって決まる値が所定値よりも大きい部分がある場合に、当該所定値よりも大きい部分が前記所定の有声音区間の始端に近い場合は前記大きい部分から前記始端までの区間について、前記大きい部分が前記所定の有声音区間の終端に近い場合は前記大きい部分から前記終端まで

の区間について前記補間を行うことを特徴とする韻律表示装置。

【請求項7】 請求項1において、前記表示手段は、前記修正手段による前記時間的対応付けの結果、前記モデル音声の有声音区間において基本周波数の時間軌跡を非表示とした部分に対応する前記入力音声の対応区間の基本周波数の時間軌跡を非表示とすることを特徴とする韻律表示装置。

【請求項8】 音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声に所定の変換処理を行なって前記モデル音声の韻律に対応させて再生出力する音声言語の再生装置であって、前記入力音声の韻律を分析する分析手段と、前記分析手段による分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正手段と、前記修正手段による修正結果に従って再生音声を生成する手段であって、前記入力音声を前記修正結果に従って時間伸縮し、前記修正手段による前記時間的対応付けの結果に従って前記入力音声の有声音区間の基本周波数に対応する前記モデル音声の有声音区間の基本周波数と入れ替え、さらに、前記入れ替えた基本周波数の時間軌跡のほぼ平均値を前記入力音声の基本周波数の時間軌跡のほぼ平均値に合わせるように時間軌跡を周波数軸でシフトした前記再生音声を生成する再生手段とを備えたことを特徴とする再生装置。

【請求項9】 請求項8において、前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込む手段と、前記修正手段による前記修正結果に平滑化処理を行なう平滑化手段とを備え、前記再生手段は、前記平滑化された前記モデル音声の基本周波数の時間軌跡と前記平滑化処理された前記入力音声の基本周波数の時間軌跡の差分を求め、前記差分と前記平滑化手段による平滑化前の前記入力音声の基本周波数とを加算することで、前記入れ替えを行なうことを特徴とする再生装置。

【請求項10】 音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声と前記モデル音声の韻律の類似度を判定する類似度判定装置であって、前記入力音声の韻律を分析する分析手段と、前記分析手段による分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正手段と、前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込む手段と、前記取り込んだ時間軌跡と前記修正手段による修正結果

に従って前記類似度の判定を行なう判定手段とを備えたことを特徴とする類似度判定装置。

【請求項11】 請求項10において、前記判定手段は、前記修正手段による修正結果に従って、前記入力音声の基本周波数の時間軌跡を算出する手段と、前記取り込んだ時間軌跡および前記算出した時間軌跡のフレーズ全体にわたる各平均値を算出する手段と、前記取り込んだ時間軌跡と前記算出した時間軌跡の差分から前記各平均値の差分を減算する手段とを備えたことを特徴とする類似度判定装置。

【請求項12】 請求項1～7のいずれかに記載の韻律表示装置と、請求項8または9に記載の再生装置とを備えたことを特徴とする音声言語処理装置。

【請求項13】 請求項1～7のいずれかに記載の韻律表示装置と、請求項10または11に記載の類似度判定装置とを備えたことを特徴とする音声言語処理装置。

【請求項14】 請求項8または9に記載の再生装置と、請求項10または11に記載の類似度判定装置とを備えたことを特徴とする音声言語処理装置。

【請求項15】 請求項1～7のいずれかに記載の韻律表示装置と、請求項8または9に記載の再生装置と、請求項10または11に記載の類似度判定装置とを備えたことを特徴とする音声言語処理装置。

【請求項16】 音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声の韻律を表示する音声言語の韻律表示方法のプログラムを記憶した記憶媒体であって、前記プログラムはコンピュータに、前記入力音声の韻律を分析する分析ステップと、前記分析ステップにおける分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正ステップと、前記修正ステップにおける修正結果に従って、前記表示をグラフィックに行なうための表示データを生成する表示ステップとを実行させることを特徴とする記録媒体。

【請求項17】 請求項16において、前記修正ステップにおいて、前記入力音声および前記モデル音声それぞれの無音区間を省いてから時間的対応付けを行うことを特徴とする記録媒体。

【請求項18】 請求項16において、前記プログラムは前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する入力ステップをさらに前記コンピュータに実行させ、前記表示ステップにおいて、前記修正結果に平滑化処理を行なうことで、前記基本周波数の時間軌跡を滑らかに変化させて前記表示を行なわせることを特徴とする記録媒体。

【請求項19】 請求項16において、

前記表示ステップにおいて、前記修正ステップにおける前記時間的対応付けの結果、前記モデル音声の有声音区間が前記分析結果の無音区間に対応すると判定された場合、当該対応区間の基本周波数の時間軌跡を周辺の有声音区間の値を用いて補間し、前記モデル音声の無音区間が前記分析結果の有声音区間に対応すると判定された場合、当該対応区間は非表示とすることを特徴とする記録媒体。

【請求項20】 請求項16において、前記表示ステップにおいて、前記分析結果の有声音区間の両端部で抽出された基本周波数が時間変化に対し不規則に変化すると前記修正ステップによって判定された場合、当該変化部分のデータを前記有声音区間の別の部分の基本周波数の時間軌跡の値を用いて補間することを特徴とする記録媒体。

【請求項21】 請求項20において、前記プログラムは前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析ステップをさらに前記コンピュータに実行させ、前記表示ステップにおいて、前記分析結果の所定の有声音区間の端部において、近傍の複数のフレームの基本周波数と前記有声音区間の長さとの前記有声音区間における時間位置によって決まる値が所定値よりも大きい部分がある場合に、当該所定値よりも大きい部分が前記所定の有声音区間の始端に近い場合は前記大きい部分から前記始端までの区間について、前記大きい部分が前記所定の有声音区間の終端に近い場合は前記大きい部分から前記終端までの区間について前記補間を行うことを特徴とする記録媒体。

【請求項22】 請求項16において、前記表示ステップにおいて、前記修正ステップにおける前記時間的対応付けの結果、前記モデル音声の有声音区間において基本周波数の時間軌跡を非表示とした部分に対応する前記入力音声の対応区間の基本周波数の時間軌跡を非表示とすることを特徴とする記録媒体。

【請求項23】 音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声に所定の変換処理を行なって前記モデル音声の韻律に対応させて再生出力する再生方法のプログラムを記憶した記憶媒体であって、

前記プログラムはコンピュータに、前記入力音声の韻律を分析する分析ステップと、前記分析ステップにおける分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正ステップと、

前記修正ステップにおける修正結果に従って再生音声を生成するステップであって、前記入力音声を前記修正結果に従って時間伸縮し、前記修正ステップにおける前記時間的対応付けの結果に従って前記入力音声の有声音区間の基本周波数を対応する前記モデル音声の有声音区間

10

20

30

40

50

の基本周波数と入れ替え、さらに、前記入れ替えた基本周波数の時間軌跡のほぼ平均値を前記入力音声の基本周波数の時間軌跡のほぼ平均値に合わせるように時間軌跡を周波数軸でシフトした前記再生音声を生成する再生ステップとを実行させることを特徴とする記録媒体。

【請求項24】 請求項23において、前記プログラムは前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込むステップと、前記修正ステップにおける前記修正結果に平滑化処理を行なう平滑化ステップとをさらに前記コンピュータに実行させ、

前記再生ステップにおいて、前記平滑化された前記モデル音声の基本周波数の時間軌跡と前記平滑化処理された前記入力音声の基本周波数の時間軌跡の差分を求め、前記差分と前記平滑化ステップにおける平滑化前の前記入力音声の基本周波数とを加算することで、前記入れ替えを行なうことを特徴とする記録媒体。

【請求項25】 音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声と前記モデル音声の韻律の類似度を判定する方法のプログラムを記憶した記憶媒体であって、

前記プログラムはコンピュータに、前記入力音声の韻律を分析する分析ステップと、前記分析ステップにおける分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正ステップと、前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込むステップと、前記取り込んだ時間軌跡と前記修正ステップにおける修正結果に従って前記類似度の判定を行なう判定ステップとを実行させることを特徴とする記録媒体。

【請求項26】 請求項25において、前記判定ステップは、前記修正ステップにおける修正結果に従って、前記入力音声の基本周波数の時間軌跡を算出するステップと、前記取り込んだ時間軌跡および前記算出した時間軌跡のフレーズ全体にわたる各平均値を算出するステップと、前記取り込んだ時間軌跡と前記算出した時間軌跡の差分から前記各平均値の差分を減算するステップとを備えたことを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は音声言語の韻律表示装置、再生装置、類似度判定装置および記録媒体に関し、特に詳細には、語学教育や音声言語の研究などにおいて、アクセント／イントネーションや個々の音の長さのバランスなどの所謂「韻律」を学習したり、分析したりする際の視覚的／聴覚的支援技術に用いられる音声言語の韻律表示装置、再生装置、類似度判定装置および記

録媒体に関する。

【0002】

【従来の技術】従来、語学教育における韻律の学習法は、教師の模範発声を聴取後、学習者がこれを真似て発声し、教師が講評するという聴覚中心の学習であった。教師が同席せず、テープなどに録音された模範音声（モデル音声）のみを聴取する場合においては、学習者は自らの発声の良し悪しを判断する手段がなかった。

【0003】そこで、近年コンピュータを使って、語学教育等における韻律の習得を支援する試み（特開平3-252700号公報、電子情報通信学会技術研究報告SP96-18など）が行われるようになってきた。このようにコンピュータを使うことによって、聴覚的のみならず、視覚的にも学習者を支援することが可能になる。これらには、音声学や音響学の分野で研究用に開発された計算機によるデジタル音声信号処理技術が応用されている。

【0004】

【発明が解決しようとする課題】しかし、この種のデジタル音声信号処理技術においては、音声の録音条件が悪い場合、例えば環境騒音が大きい、マイクの使い方が不適切、発声の仕方が悪いなどの条件下では正しい処理が行えず、基本周波数の誤抽出などの誤動作が生じやすい。また、誤動作しない場合でも、研究用の詳細な分析結果をそのまま文字やグラフにしてコンピュータ画面に表示しても、そのユーザーである専門知識のない一般人にとっては、語学学習などの参考として分かりにくいなどの欠点があり、一般ユーザーにも分かりやすく提示する技術は現在まで実用化されていない。

【0005】本発明は上記の事情に鑑み、語学教育や音声言語の研究などにおいて、アクセント／イントネーションや個々の音の長さのバランスなどの所謂「韻律」を対象とした場合に、予めモデルとなる模範的な発声をコンピュータの記憶媒体に記憶しておき、これと同一内容をユーザーが発声した場合に、その韻律の特徴をユーザー音声とモデル音声とを比較して視覚的に表示したり、ユーザーの音声の韻律をモデル音声の韻律にそっくり入れ替えた変換音声を作成して聴覚的に呈示することで、ユーザーへの語学教育や音声言語の研究を支援することのできる音声言語の韻律表示装置、再生装置、類似度判定装置および記録媒体を提供することを目的としている。

【0006】

【課題を解決するための手段】上記の目的を達成するために請求項1の発明は、音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声の韻律を表示する音声言語の韻律表示装置であって、前記入力音声の韻律を分析する分析手段と、前記分析手段による分析結果と前記モデル音声との時間的

対応付けを行なうことで、前記分析結果を修正する修正手段と、前記修正手段による修正結果に従って、前記表示をグラフ的行なうための表示データを生成する表示手段とを備えた形態の韻律表示装置を実施した。

【0007】請求項2の発明は、請求項1において、前記修正手段は、前記入力音声および前記モデル音声それぞれの無音区間を省いてから時間的対応付けを行う形態の韻律表示装置を実施した。

【0008】請求項3の発明は、請求項1において、前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析手段を備え、前記表示手段は、前記修正結果に平滑化処理を行なうことで、前記基本周波数の時間軌跡を滑らかに変化させて前記表示を行なわせる形態の韻律表示装置を実施した。

【0009】請求項4の発明は、請求項1において、前記表示手段は、前記修正手段による前記時間的対応付けの結果、前記モデル音声の有声部が前記分析結果の無声部に対応すると判定された場合、当該対応区間の基本周波数の時間軌跡を周辺の有声部の値を用いて補間し、前記モデル音声の無声部が前記分析結果の有声部に対応すると判定された場合、当該対応区間は非表示とする形態の韻律表示装置を実施した。

【0010】請求項5の発明は、請求項1において、前記表示手段は、前記分析結果の有声音区間の両端部で抽出された基本周波数が時間変化に対し不規則に変化すると前記修正手段によって判定された場合、当該変化部分のデータを前記有声音区間の別の部分の基本周波数の時間軌跡の値を用いて補間する形態の韻律表示装置を実施した。

【0011】請求項6の発明は、請求項5において、前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析手段を備え、前記表示手段は、前記分析結果の所定の有声音区間の端部において、近傍の複数のフレームの基本周波数と前記有声音区間の長さとの前記有声音区間における時間位置によって決まる値が所定値よりも大きい部分がある場合に、当該所定値よりも大きい部分が前記所定の有声音区間の始端に近い場合は前記大きい部分から前記始端までの区間について、前記大きい部分が前記所定の有声音区間の終端に近い場合は前記大きい部分から前記終端までの区間について前記補間を行う形態の韻律表示装置を実施した。

【0012】請求項7の発明は、請求項1において、前記表示手段は、前記修正手段による前記時間的対応付けの結果、前記モデル音声の有声音区間において基本周波数の時間軌跡を非表示とした部分に対応する前記入力音声の対応区間の基本周波数の時間軌跡を非表示とする形態の韻律表示装置を実施した。

【0013】請求項8の発明は、音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を

取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声に所定の変換処理を行なって前記モデル音声の韻律に対応させて再生出力する音声言語の再生装置であって、前記入力音声の韻律を分析する分析手段と、前記分析手段による分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正手段と、前記修正手段による修正結果に従って再生音声を生成する手段であって、前記入力音声を前記修正結果に従って時間伸縮し、前記修正手段による前記時間的対応付けの結果に従って前記入力音声の有声音区間の基本周波数を対応する前記モデル音声の有声音区間の基本周波数と入れ替え、さらに、前記入れ替えた基本周波数の時間軌跡のほぼ平均値を前記入力音声の基本周波数の時間軌跡のほぼ平均値に合わせるように時間軌跡を周波数軸でシフトした前記再生音声を生成する再生手段とを備えた形態の再生装置を実施した。

【0014】請求項9の発明は、請求項8において、前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込む手段と、前記修正手段による前記修正結果に平滑化処理を行なう平滑化手段とを備え、前記再生手段は、前記平滑化された前記モデル音声の基本周波数の時間軌跡と前記平滑化処理された前記入力音声の基本周波数の時間軌跡の差分を求め、前記差分と前記平滑化手段による平滑化前の前記入力音声の基本周波数とを加算することで、前記入れ替えを行なう形態の再生装置を実施した。

【0015】請求項10の発明は、音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声と前記モデル音声の韻律の類似度を判定する類似度判定装置であって、前記入力音声の韻律を分析する分析手段と、前記分析手段による分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正手段と、前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込む手段と、前記取り込んだ時間軌跡と前記修正手段による修正結果に従って前記類似度の判定を行なう判定手段とを備えた形態の類似度判定装置を実施した。

【0016】請求項11の発明は、請求項10において、前記判定手段は、前記修正手段による修正結果に従って、前記入力音声の基本周波数の時間軌跡を算出する手段と、前記取り込んだ時間軌跡および前記算出した時間軌跡のフレーズ全体にわたる各平均値を算出する手段と、前記取り込んだ時間軌跡と前記算出した時間軌跡の差分から前記各平均値の差分を減算する手段とを備えた形態の類似度判定装置を実施した。

【0017】請求項12の発明は、請求項1～7のいずれかに記載の韻律表示装置と、請求項8または9に記載の再生装置とを備えた形態の音声言語処理装置を実施し

た。

【0018】請求項13の発明は、請求項1～7のいずれかに記載の韻律表示装置と、請求項10または11に記載の類似度判定装置とを備えた形態の音声言語処理装置を実施した。

【0019】請求項14の発明は、請求項8または9に記載の再生装置と、請求項10または11に記載の類似度判定装置とを備えた形態の音声言語処理装置を実施した。

【0020】請求項15の発明は、請求項1～7のいずれかに記載の韻律表示装置と、請求項8または9に記載の再生装置と、請求項10または11に記載の類似度判定装置とを備えた形態の音声言語処理装置を実施した。

【0021】請求項16の発明は、音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声の韻律を表示する音声言語の韻律表示方法のプログラムを記憶した記憶媒体であって、前記プログラムはコンピュータに、前記入力音声の韻律を分析する分析ステップと、前記分析ステップにおける分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正ステップと、前記修正ステップにおける修正結果に従って、前記表示をグラフィックに行なうための表示データを生成する表示ステップとを実行させる形態の記録媒体を実施した。

【0022】請求項17の発明は、請求項16において、前記修正ステップにおいて、前記入力音声および前記モデル音声それぞれの無音区間を省いてから時間的対応付けを行う形態の記録媒体を実施した。

【0023】請求項18の発明は、請求項16において、前記プログラムは前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析ステップをさらに前記コンピュータに実行させ、前記表示ステップにおいて、前記修正結果に平滑化処理を行なうことで、前記基本周波数の時間軌跡を滑らかに変化させて前記表示を行なわせる形態の記録媒体を実施した。

【0024】請求項19の発明は、請求項16において、前記表示ステップにおいて、前記修正ステップにおける前記時間的対応付けの結果、前記モデル音声の有声部が前記分析結果の無声部に対応すると判定された場合、当該対応区間の基本周波数の時間軌跡を周辺の有声部の値を用いて補間し、前記モデル音声の無声部が前記分析結果の有声部に対応すると判定された場合、当該対応区間は非表示とする形態の記録媒体を実施した。

【0025】請求項20の発明は、請求項16において、前記表示ステップにおいて、前記分析結果の有声音区間の両端部で抽出された基本周波数が時間変化に対し、不規則に変化すると前記修正ステップによって判定された場合、当該変化部分のデータを前記有声音区間の別の

部分の基本周波数の時間軌跡の値を用いて補間する形態の記録媒体を実施した。

【0026】請求項21の発明は、請求項20において、前記プログラムは前記入力音声の有声音区間についての基本周波数を所定時間間隔のフレーム単位で抽出する他の分析ステップをさらに前記コンピュータに実行させ、前記表示ステップにおいて、前記分析結果の所定の有声音区間の端部において、近傍の複数のフレームの基本周波数と前記有声音区間の長さとの前記有声音区間における時間位置によって決まる値が所定値よりも大きい部分がある場合に、当該所定値よりも大きい部分が前記所定の有声音区間の始端に近い場合は前記大きい部分から前記始端までの区間について、前記大きい部分が前記所定の有声音区間の終端に近い場合は前記大きい部分から前記終端までの区間について前記補間を行う形態の記録媒体を実施した。

【0027】請求項22の発明は、請求項16において、前記表示ステップにおいて、前記修正ステップにおける前記時間的対応付けの結果、前記モデル音声の有声音区間において基本周波数の時間軌跡を非表示とした部分に対応する前記入力音声の対応区間の基本周波数の時間軌跡を非表示とする形態の記録媒体を実施した。

【0028】請求項23の発明は、音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声に所定の変換処理を行なって前記モデル音声の韻律に対応させて再生出力する再生方法のプログラムを記憶した記憶媒体であって、前記プログラムはコンピュータに、前記入力音声の韻律を分析する分析ステップと、前記分析ステップにおける分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正ステップと、前記修正ステップにおける修正結果に従って再生音声を生成するステップであって、前記入力音声を前記修正結果に従って時間伸縮し、前記修正ステップにおける前記時間的対応付けの結果に従って前記入力音声の有声音区間の基本周波数に対応する前記モデル音声の有声音区間の基本周波数と入れ替え、さらに、前記入れ替えた基本周波数の時間軌跡のほぼ平均値を前記入力音声の基本周波数の時間軌跡のほぼ平均値に合わせるように時間軌跡を周波数軸でシフトした前記再生音声を生成する再生ステップとを実行させる形態の記録媒体を実施した。

【0029】請求項24の発明は、請求項23において、前記プログラムは前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込むステップと、前記修正ステップにおける前記修正結果に平滑化処理を行なう平滑化ステップとをさらに前記コンピュータに実行させ、前記再生ステップにおいて、前記平滑化された前記モデル音声の基本周波数の時間軌跡と前記平滑化処理された前記入力音声の基本周波数の時間軌

跡の差分を求め、前記差分と前記平滑化ステップにおける平滑化前の前記入力音声の基本周波数とを加算することで、前記入れ替えを行なう形態の記録媒体を実施した。

【0030】請求項25の発明は、音声言語の韻律の特徴を予め分析して記憶したデータベースからモデル音声を取り込むとともに前記モデル音声と同一内容の音声を入力し、当該入力音声と前記モデル音声の韻律の類似度を判定する方法のプログラムを記憶した記憶媒体であって、前記プログラムはコンピュータに、前記入力音声の韻律を分析する分析ステップと、前記分析ステップにおける分析結果と前記モデル音声との時間的対応付けを行なうことで、前記分析結果を修正する修正ステップと、前記データベースから平滑化された前記モデル音声の基本周波数の時間軌跡を取り込むステップと、前記取り込んだ時間軌跡と前記修正ステップにおける修正結果に従って前記類似度の判定を行なう判定ステップとを実行させる形態の記録媒体を実施した。

【0031】請求項26の発明は、請求項25において、前記判定ステップは、前記修正ステップにおける修正結果に従って、前記入力音声の基本周波数の時間軌跡を算出するステップと、前記取り込んだ時間軌跡および前記算出した時間軌跡のフレーズ全体にわたる各平均値を算出するステップと、前記取り込んだ時間軌跡と前記算出した時間軌跡の差分から前記各平均値の差分を減算するステップとを備えた形態の記録媒体を実施した。

【0032】上記請求項2または17の形態によれば、入力音声を発声する学習者または研究者がフレーズ内でモデル音声には無い間を入れたり、逆に間を挿入すべきところを連続に発声した場合にもスペクトルの不整合を回避して正しい時間的対応付けを行うことができる。

【0033】上記請求項3または18の形態によれば、学習すべき韻律の特徴を単純化し、聴感的な音の高さの変化に対応して分かりやすく表示することができる。

【0034】上記請求項4～6または19～21の形態によれば、学習者または研究者による入力音声の分析結果を修正して基本周波数の時間軌跡を表示することができる。

【0035】上記請求項7または22の形態によれば、予めモデル音声の分析結果に対して、学習すべき韻律の特徴を分かりやすくするための修正を加えておくことにより、学習者または研究者の入力音声についてもそれに準拠した修正を加えた基本周波数の時間軌跡が表示できることになる。

【0036】上記請求項8または23の形態によれば、韻律の学習を聴覚的に支援することができ、例えばモデル音声は男性の声で、学習者または研究者が女性であっても、正しい女性の声の高さで模範音声の韻律を持った音声再生することができる。

【0037】上記請求項9または24の形態によれば、

学習者または研究者の入力音声の基本周波数の周波数変換を行うことができ、学習者または研究者の入力音声の基本周波数の変化量が部分的に極端に大きくなることを防止でき、基本周波数変換を高品質に行うことができる。

【0038】上記請求項10または25の形態によれば、モデル音声に対する学習者の入力音声の韻律の類似度を客観的に算出し、学習者に自らの発声のうまさを知らせることができる。

10 【0039】上記各請求項の形態によれば、語学教育や音声言語の研究などにおいて、アクセント／イントネーションや個々の音の長さのバランスなどの所謂「韻律」のモデルとなる模範的な発声と同じ内容を学習者または研究者が発声した場合に、その韻律の特徴を学習者または研究者の音声とモデル音声とを比較して視覚的に表示したり、学習者または研究者の音声の韻律をモデル音声の韻律にそっくり入れ替えた変換音声を作成して聴覚的に呈示する際に、基本周波数の誤抽出などの誤動作を低減したり、語学学習などの参考になりやすい形に分析結果を整形して表示することにより、学習者への語学教育や研究者の音声言語の研究を支援することを可能とする。

【0040】

【発明の実施の形態】図1は本発明に係る音声言語の韻律表示装置、再生装置、類似度判定装置、および音声言語処理装置の一実施例を示すブロック図である。

30 【0041】100は上記各装置として動作するコンピュータ装置であり、CPU、主記憶装置としてのROM、補助記憶装置としてのHDD、FDD、CPUの作業エリアおよび一時的なデータ記憶領域としてのRAM、外部との音声データのやり取りに用いられるI/F等のハードウェアを備えた周知の構成の汎用コンピュータ、例えば、市販のパーソナルコンピュータやワークステーションを使用することができる。これらハードウェアは図示を省略し、CPUによる処理の流れを機能的ブロック図として示した。また、101はハードディスクアレー等のモデル音声データベース部、120は表示装置である。

40 【0042】この図に示すコンピュータ装置100は、上記補助記憶装置または外部記憶装置（図示せず）等の本発明に係る記録媒体から音声言語の韻律表示方法、再生方法、類似度判定方法、または音声言語処理方法のプログラムをロードし、A/D変換部102、分析部103、時間的対応付け部104、修正部105、時間伸縮部106、平滑化／補間部107、表示データ生成部108、差分算出部109、加算部110、基本周波数変換部111、D/A変換部112、および類似度判定部113としての機能を遂行することができる。

50 【0043】これら機能により、語学教育や音声言語の研究などにおいて、アクセント／イントネーションや個

々の音の長さのバランスなどの所謂「韻律」を対象とした場合に、予めモデルとなる模範的な発声がディスク記憶装置等に用意されていることを前提として、これと同じ内容をユーザー（学習者または研究者）が発声した場合に、その韻律の特徴をユーザー音声とモデル音声とを比較して視覚的に表示したり、ユーザーの音声の韻律をモデル音声の韻律にそっくり入れ替えた変換音声を作成して聴覚的に呈示する際に、基本周波数の誤抽出などの誤動作を低減したり、語学学習などの参考になりやすい形に分析結果を整形して表示することにより、ユーザーへの語学教育や音声言語の研究を支援する手段を提供するものである。

【0044】図1におけるモデル音声データベース部101は、例えば語学学習の場合、学習の基本となるフレーズを予めネイティブなアナウンサーなどが模範音声として発声したもの（以後、“モデル音声”と呼ぶ）を、その基本周波数およびパワーやルビの位置、及びその音声波形などとともにデータベース化してハードディスク等の記憶装置（図示せず）に記憶したものである。

【0045】データベース化の際には、“モデル音声”に対して、例えばフレーム周期5msで有声／無声／無音の判定と基本周波数の抽出（例えば、電子情報通信学会論文誌A、Vol. J80-ANo. 9, pp. 1341-1350などの方法による）を行い、さらに手修正を加えた後、各有声音区間単位で基本周波数の平滑化（例えば、カットオフ周波数 $f_c = 6\text{Hz}$ のLPF）を行って基本周波数の時間軌跡を求める。この際、例えば中国語では音節毎に第1声～第4声および轻声のいずれかの音調がつくことを理解しやすくするため、実際には有声音が連続する場合でも、基本周波数の時間軌跡上で各音節の境界部分は非表示とし、切れ目を入れる修正を目視で加えてデータベース化することにより、学習者または研究者の理解を促進することができる。

【0046】上記コンピュータ装置100の処理を概略的に説明すると、まず、ユーザーがD/A変換部112およびスピーカ（図示せず）を介してモデル音声を聴取したり、表示装置120の画面に表示されたモデル音声の基本周波数の時間軌跡を参考にして、モデル音声を真似て、コンピュータ装置100に接続されたマイクロホン（図示せず）に向かって発声（以後、“ユーザー音声”と呼ぶ）すると、A/D変換部102を介して“ユーザー音声”が音声波形としてコンピュータ装置100内に取り込まれる。取り込まれた音声波形は、その基本周波数の時間軌跡が自動的に表示装置120の画面に表示され、ユーザーは、自分が発声したものとモデル音声の基本周波数の時間軌跡との違いを視覚的に知ることができる。

【0047】同時に、基本周波数変換部111により、ユーザー音声の音質を保ちながらモデル音声の韻律を持った、韻律変換された変換音声を生成する。したがって

ユーザーは、スピーカからの再生音によりユーザー音声とモデル音声を聞き比べるだけでなく、同じユーザー音声の音質の韻律変換前と韻律変換後の音声を聴覚的に聞き比べることもできる。さらに、類似度判定部113により、モデル音声に対するユーザー音声の基本周波数時間軌跡に関する類似度が判定され、判定結果が表示データ生成部108に出力され、表示データに従って表示装置120の画面に表示される。上記モデル音声データベース101とコンピュータ装置100を用いた本実施形態によれば、ユーザーに対し視覚的および聴覚的支援を実現して学習効果を高めることができる。

【0048】以下、ユーザー音声の基本周波数時間軌跡を表示する処理手順について説明する。コンピュータ装置100に取り込まれてA/D変換されたユーザー音声波形に対し、分析部103により、例えばフレーム周期5msで有声／無声／無音が判定され、基本周波数が抽出（例えば、電子情報通信学会論文誌A、J80-ANo. 9, pp. 1341-1350などの方法による）される。

【0049】分析部103による上記処理の一方で、モデル音声データベース部101に予め記憶してある、モデル音声の音声波形、有声／無声／無音の情報、有声音区間内の非表示位置情報が時間的対応付け部104に取り込まれる。さらに時間的対応付け部104には、ユーザー音声波形と、分析部103により得られた有声／無声／無音の情報が取り込まれる。

【0050】次に、時間的対応付け部104はこれらの情報を基に、モデル音声とユーザー音声の間で、LPCケプストラムによるDPマッチングを用いて、両者の音声波形の時間的対応付けを行なう。ところで、ユーザーは、フレーズ内でモデル音声には無い間を入れたり、逆に間を挿入すべきところを連続して発声することがある。このような発声に対してはスペクトル的に不整合が生じるため、時間的対応付けが正しく行われないことが多い。そこで、両者の音声の無音と判定されている区間を省いて（削除して）からDPマッチングを行って、両者の音声間の有声／無声の照合結果および有声音区間内の非表示位置の照合結果を生成する。時間的対応付け部104はまた、モデル音声に対するユーザー音声の時間伸縮情報を生成する。

【0051】修正部105は、上記の照合結果と、分析部103によって一定の時間間隔でフレーム毎に得たユーザー音声の基本周波数とに従い以下の処理を行う。

【0052】修正部105によってまず、一定の時間間隔のフレーム毎に得られたユーザー音声の基本周波数について、一連の有声音区間の両端のいくつかのフレームで抽出された基本周波数が乱れ、時間変化に対して不規則に変化している場合、その値は後に行なう平滑化／補間部107による平滑化のデータに含めないことにする。乱れていない場合には、滑らかに変化する基本周波

数の時間軌跡を表示するために、平滑化／補間部107による平滑化処理をう。

【0053】基本周波数の乱れの判定は、具体的には次＊

$$A = W_r \cdot \sqrt{\frac{|1.0 - P_n / P_{n-1}|}{C_1}} \quad (i)$$

ここで

$$W_r = \begin{cases} C_2 \cdot \left(1.0 - \frac{D}{T}\right) & (D < T) \\ 0 & (D \geq T) \end{cases}$$

P_n (Hz): 当該フレームの基本周波数

P_{n-1} (Hz): P_n の1フレーム内側の基本周波数

D (ms): 当該フレームの有声区間の開始(終了)時刻からの時間

T (ms): $L/2$ L は有声音区間長 但し $L > 2 \cdot C_3$ (ms)の場合 $T = C_3$ (ms)

C_1 は定数で、 $0 < C_1 < 1.0$

C_2 は定数で、 $1.0 \leq C_2 \leq 5.0$

C_3 は定数で、 $20 \leq C_3 \leq 500$

【0055】式(1)において、値Aが1以上になった場合には、以下のように平滑化／補間部107による補間の対象区間を決定する。

【0056】例えば、図2に示すように有声音区間の終端(終了点)側において式(1)の値Aが1以上になった場合、このフレームより後の部分のフレーム(Aが1以上になったフレームを含む)のデータは平滑化のデータから除外し、除外したデータを、後に平滑化／補間部107によって補間の対象区間とする。図2は有声音区間の終端側について示したが、有声音区間の始端(開始点)側において式(1)の値Aが1以上になった場合は、このフレームより前の部分のフレーム(Aが1以上になったフレームを含む)のデータは平滑化のデータから除外し、除外したデータを、後に平滑化／補間部107によって補間の対象区間とする。

【0057】なお、式(1)中の定数 $C_1 \sim C_3$ として、 $C_1 = 0.2$ 、 $C_2 = 3.0$ 、 $C_3 = 200$ 程度を用いると良い結果が得られる。

【0058】続いて修正部105により、時間的対応付け部104によるDPマッチングの結果、モデル音声の有声部にユーザー音声の無声部が対応した場合、そのユーザー音声の区間についても、後の平滑化／補間部107による基本周波数の平滑化のデータから除外し、平滑化／補間部107による補間の対象区間とする。一方、時間的対応付け部104によるDPマッチングの結果、モデル音声の無声部にユーザー音声の有声部が対応した場合、その区間は基本周波数時間軌跡を非表示とするように修正する。

【0059】さらに続いて修正部105により、時間的対応付け部104によるDPマッチングの結果、モデル音声の有声音区間でも基本周波数時間軌跡を非表示とした部分に対応するユーザー音声の区間について、同様に基本周波数時間軌跡を非表示とするように修正する。

【0060】次に、上記の通りに修正されたユーザー音

＊式(1)に従って行なう。

【0054】

【数1】

声の基本周波数は、時間伸縮部106により、前述の通り時間的対応付け部104によって生成された時間伸縮情報に従ってモデル音声に合わせ時間伸縮される。時間伸縮された基本周波数データ $P_o(t)$ に対しては、平滑化／補間部107によって、有声音区間で、非表示区間にも補間対象区間にもなっていない部分について平滑化(例えば、カットオフ周波数 $f_c = 6\text{Hz}$ のLPF)処理を行う。さらに平滑化／補間部107によって、この平滑化データを用い、修正部105によって前述の通りに補間対象区間とされている部分をその周辺の平滑化済み基本周波数時間軌跡の値から最小自乗法により補間し、ユーザー音声の平滑化済み基本周波数時間軌跡 $P_u(t)$ を生成する。

【0061】最後に表示データ生成部108によって、モデル音声の平滑化済み基本周波数時間軌跡の平均値とユーザー音声の平滑化済み基本周波数時間軌跡の平均値を合わせるように、基本周波数時間軌跡を周波数軸方向に適宜シフトさせて調整し、上記の通り整形された時間軌跡を表示装置120の画面に表示させる。図3は実際の表示の一例を示し、白い軌跡がモデル音声の平滑化済み基本周波数時間軌跡、黒い軌跡がユーザー音声の平滑化済み基本周波数時間軌跡であり、これら軌跡とともに、画面上部に対応する中国語の文字の表示とルビ(読み方)の表示がなされている。

【0062】次に、基本周波数変換部111等による基本周波数変換音声作成の処理手順について説明する。この処理は、ユーザー音声の平滑化済み基本周波数時間軌跡 $P_u(t)$ が得られていれば、上記の表示処理と平行して実行される。

【0063】まず、差分算出部109によりモデル音声データベース部101からモデル音声の平滑化済み基本周波数時間軌跡 $P_m(t)$ を取得し、平滑化／補間部107によるユーザー音声の平滑化済み基本周波数時間軌跡 $P_u(t)$ との差分($P_m(t) - P_u(t)$)を求

める。しかしこの差分には、両者の平均的な声の高さの差も含まれている。そこで、差分算出部109によりさらに、 $P_m(t)$ 、 $P_u(t)$ それぞれのフレーズ全体にわたる平均値 P_{mav} 、 P_{uav} を求め、この差分を上記差分から減算してキャンセルすることで、純粋な基本周波数の時間変化の差分 $P_d(t) = (P_m(t) - P_u(t)) - (P_{mav} - P_{uav})$ を算出する。

【0064】次に、加算部110により、差分算出部109によって算出された差分 $P_d(t)$ と平滑化される前のユーザー音声の基本周波数 $P_o(t)$ とを加算して、最終的な変換音声の基本周波数の時間変化 $P_o(t) + P_d(t)$ を得る。これにより、ユーザー音声の有声音区間の基本周波数を、時間的対応付けの結果により対応関係にあるモデル音声の有声音区間の基本周波数に入れ替えた合成音声を作成されたことになる。

【0065】このようにして、変換音声の基本周波数の時間変化、つまり、変換音声の基本周波数を求めることにより、例えばモデル音声が男性の声でユーザー音声は女性の声であっても、ユーザーの女性の声の高さで模範音声の韻律を持った音声を再生できるだけでなく、ユーザー音声の基本周波数の変化量が部分的に極端に大きくなることを防止でき、基本周波数変換を高品質に行うことができる。

【0066】すなわち、基本周波数変換部111（例えば、電子情報通信学会論文誌A、Vol. J73-A No. 3, pp. 387-396などの方法による）により、A/D変換したユーザー音声波形の基本周期の変換（モデル音声の有声音区間の基本周波数に入れ替えた基本周波数の時間軌跡の平均的な値をユーザー音声の基本周波数の時間軌跡の平均的な値に合わせるように時間軌跡全体を周波数軸でシフトさせる）や個々の音の継続時間長の変換を加算部110による基本周波数の時間変化 $P_o(t) + P_d(t)$ に従って逐次行ない、韻律をモデル音声のものに入れ替えた変換音声を生じ、D/A変換した音声をスピーカ（図示せず）から出力する。

【0067】次に、モデル音声に対するユーザー音声の基本周波数時間軌跡に関する類似度判定の手順について説明する。この類似度判定処理は、上記基本周波数変換音声作成処理と並列して行なうことができる。

【0068】差分算出部109により差分 $P_d(t)$ が求められると、類似度判定部113により、この差分 $P_d(t) = (P_m(t) - P_u(t)) - (P_{mav} - P_{uav})$ について、例えばその自乗を次式(2)のようにフレーズ区間全体にわたって積分して得られる値Bを基準として、韻律の類似度を判定する。但し、(2)式においてTはフレーズの時間長であり、無声区間や無音区間、あるいは修正部105により非表示と判定された区

間など、基本周波数の時間軌跡が表示されない部分は $P_d(t) = 0$ として計算する。

【0069】

【数2】

$$B = \int_0^T P_d^2(t) dt \quad (2)$$

【0070】(2)式に従い求められた類似度の判定結果Bは、数値またはグラフなどによって、表示装置120の画面に表示される（図示せず）。

10 【0071】

【発明の効果】以上説明したように本発明に係る音声言語の韻律表示装置、再生装置、類似度判定装置、音声言語処理装置、および記録媒体によれば、語学教育や音声言語の研究などにおいて、アクセント/イントネーションや個々の音の長さのバランスなどの所謂「韻律」のモデルとなる模範的な発声と同じ内容をユーザーが発声した場合に、その韻律の特徴をユーザー音声（学習者または研究者の入力音声）とモデル音声とを比較して視覚的に表示したり、ユーザー音声の韻律をモデル音声の韻律にそっくり入れ替えた変換音声を作成して聴覚的に呈示する際に、基本周波数の誤抽出などの誤動作を低減したり、語学学習などの参考になりやすい形に分析結果を整形して表示することにより、学習者への語学教育や研究者の音声言語の研究を支援することができる。

【図面の簡単な説明】

【図1】本発明に係る音声言語の韻律表示装置、再生装置、類似度判定装置、および音声言語処理装置の一実施例を示すブロック図である。

【図2】本発明による有声音区間の終端側において式(1)の値Aが1以上になった場合の例を示す説明図である。

【図3】本発明に係る一実施例による表示装置の実際の表示例を示す説明図である。

【符号の説明】

- 101 モデル音声データベース部
- 102 A/D変換部
- 103 分析部
- 104 時間的対応付け部
- 105 修正部
- 106 時間伸縮部
- 107 平滑化/補間部
- 108 表示データ生成部
- 109 差分算出部
- 110 加算部
- 111 基本周波数変換部
- 112 D/A変換部
- 113 類似度判定部

40

Fig. /

【図1】

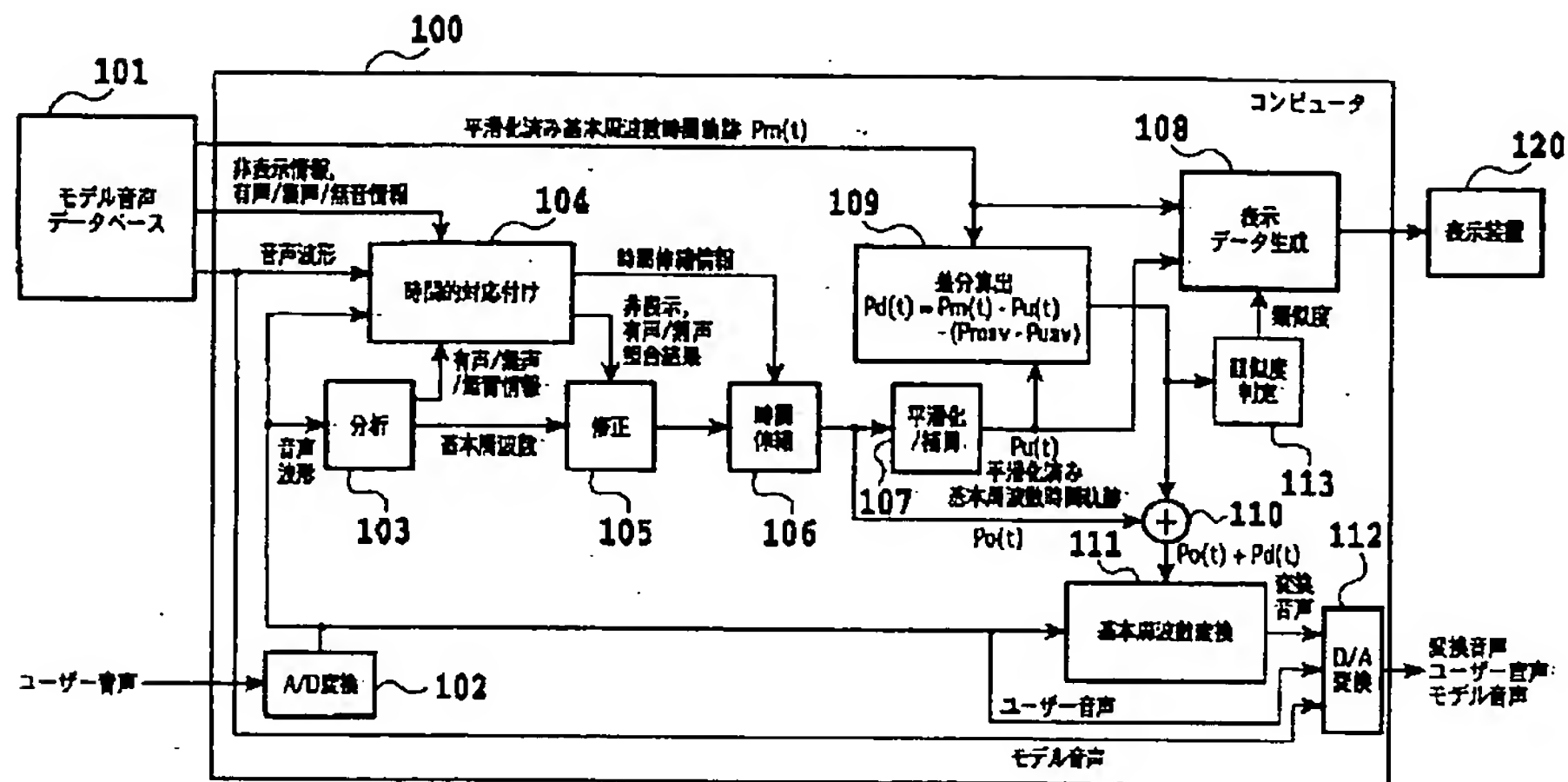


Fig. 2

【図2】

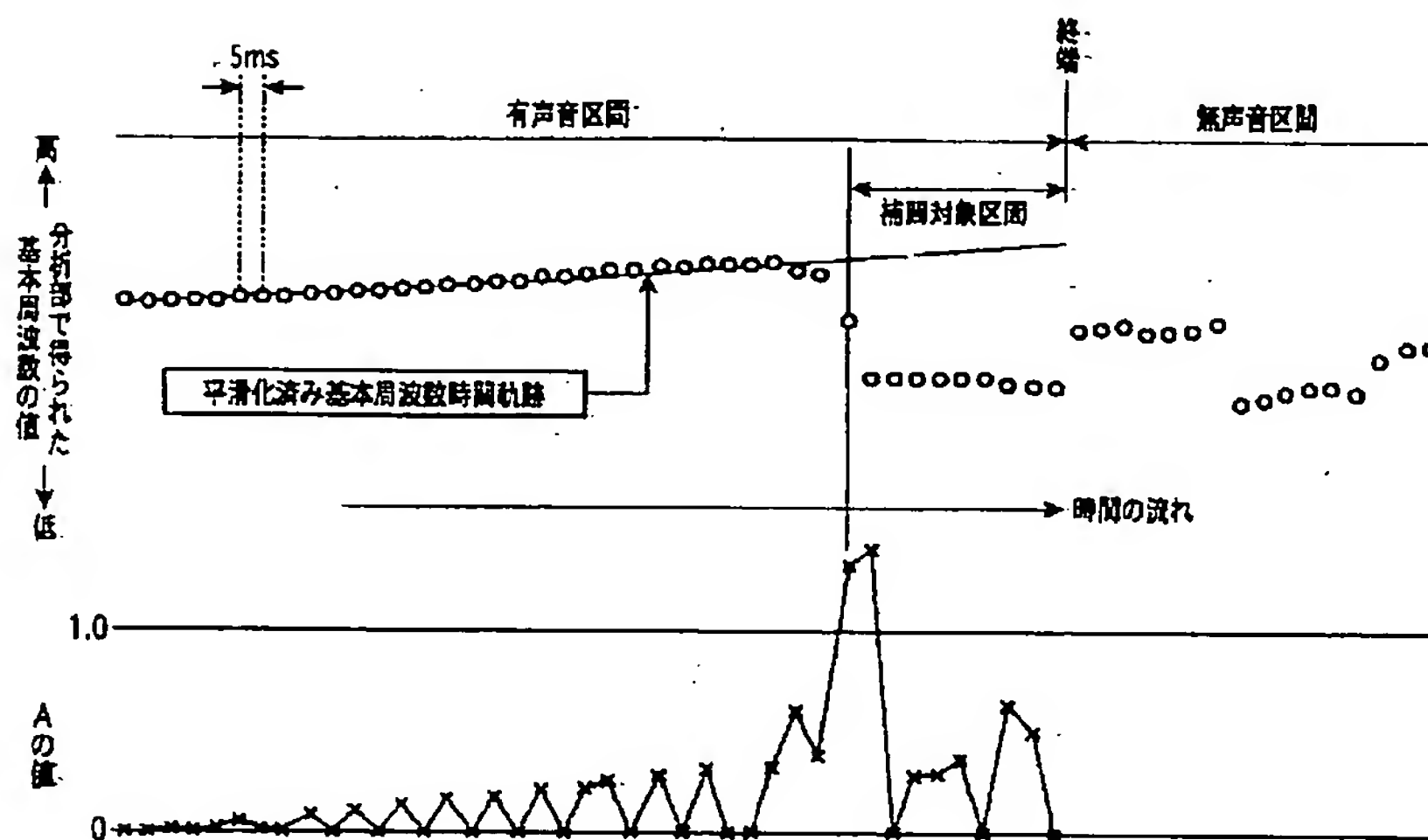
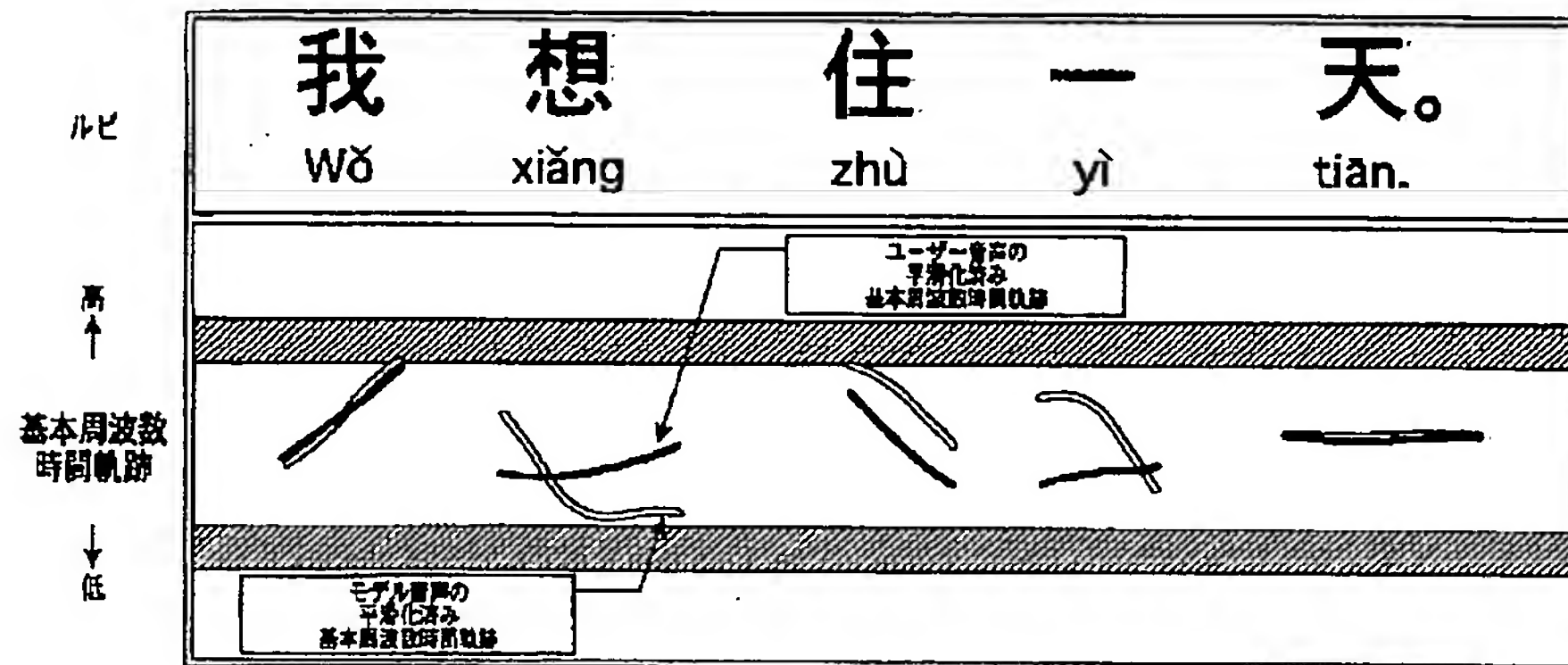


Fig. 3

〔図3〕



フロントページの続き

(51)Int.Cl. ⁷	識別記号	F I	テマコード (参考)
G 0 9 B	19/00	G 0 9 B 19/06	
G 1 0 L	11/04	G 1 0 L 3/00	S
	11/02		C
	13/08		B
	15/00		H
	15/02		5 5 1 E
// G 0 9 B	19/04	9/00	3 0 1 A
	19/06		

F ターム (参考) 2C028 AA00
 5B049 BB21 CC02 CC31 DD03 DD05
 EE03 EE05 EE07 FF03 FF04
 FF06 FF09
 5B075 ND14 NS02 PP07 PP12 PP30
 PQ02 PQ04 PQ14 PR06 QM08
 UU40
 5D015 CC12 CC13 KK02 LL05
 5D045 AB13 AB30

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.